DOCUMENT RESUME

ED 299 295 TM 012 078

AUTHOR Stephenson, Robert S.; Giacoboni, Kathryn N.

TITLE A Comparison of 1987 Results of SSAT-I Writing and Production Writing Assessment.

INSTITUTION Dade County Public Schools, Miami, FL. Office of Educational Accountability.

PUB DATE Jan 88
NOTE 46p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Capitalization (Alphabetic); *Essay Tests; Grade 10; Grammar; High School Students; *Multiple Choice Tests; Punctuation; Secondary Education; Sentence Structure; Spelling; Test Reliability; *Test Validity; *Writing Evaluation
IDENTIFIERS *Florida State Student Assessment Test; *Production Writing Assessment

ABSTRACT

This study compares the results of an indirect measure of writing ability, the Florida State Student Assessment Test, Part I (SSAT-I), and the Production Writing Assessment (PWA). Both instruments were administered to 10th graders in Dade County (Florida) during the spring of 1987. The SSAT-I is an untimed test containing a section indirectly measuring writing ability through multiple-choice items that address specific skills such as punctuation, capitalization, spelling, and grammar. The PWA directly measures writing performance by requiring students to write an essay on a given topic within one class period. A holistic scoring method (for 3,386 papers) and an analytical scoring method were used in the PWA. Results show that: (1) SSAT-I results agree with those from analytically scored production writing only when students have fairly high levels of overall writing skill; (2) scores representing the same skills for below-average students tend to be lower on the SSAT-I than on the analytically scored papers; and (3) SSAT-I results tend to penalize students with lower levels of overall writing skills. It appears that use of the writing portion of the SSAT-I as the criterion for labeling schools deficient is, at best, debatable. Examples of holistically scored students' essays from the 1986 PWA and results of analytical scoring of the PWA are appended. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299295

DCPS

DADE COUNTY PUBLIC SCHOOLS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A COMPARISON OF 1987 RESULTS OF SSAT-I WRITING AND PRODUCTION WRITING ASSESSMENT

January 1988

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert Stephenson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IM 012 078

OFFICE OF EDUCATIONAL ACCOUNTABILITY

Dade County Public Schools
Office of Educational Accountability
1450 Northeast Second Avenue
Miami, Florida 33132

A COMPARISON OF 1987 RESULTS OF SSAT-I WRITING AND PRODUCTION WRITING ASSESSMENT

January 1988

Authors: Dr. Robert S. Stephenson
and Kathryn N. Giacoboni
Testing Department

THE SCHOOL BOARD OF DADE COUNTY, FLORIDA

Dr. Michael Krop, Chairman
Mr. G. Holmes Braddock, Vice-Chairman
Mr. Paul L. Cejas
Dr. Rosa Castro Feinberg
Ms. Janet R. McAliley
Mr. Robert Renick
Mr. William H. Turner

Dr. Joseph A. Fernandez
Superintendent of Schools

TABLE OF CONTENTS

Page

EXECUTIVE SUMMARY ii

PURPOSE OF STUDY 1

DESCRIPTION OF ASSESSMENTS UNDER COMPARISON 1

OUTCOMES EXPECTED OF THE STUDY 2

PROCEDURES OF THE STUDY 2

RESULTS OF THE STUDY 5

CHART: PERCENT PASSING WRITING SKILLS IN SSAT–I AND PRODUCTION WRITING ASSESSMENT 6

CONCLUSIONS OF THE STUDY 7

APPENDIX A: SSAT–I Writing Skills Assessed in 1987 9

APPENDIX B, PART I: Holistic Scoring Guide 11

APPENDIX C, PART II Examples of Holistically Scored Students' Essays from the 1986 Production Writing Test 13

APPENDIX C: Results of Analytical Scoring: Comparisons of Results for Skills Where Greatest Data Differences Occur 33

EXECUTIVE SUMMARY

This study compares the results of an indirect and a direct measure of writing ability: the Florida State Student Assessment Test, Part I (SSAT–I), and the Production Writing Assessment, both administered to Dade County tenth graders in the spring of 1987. The purpose of this study is to determine similarities and/or differences between results of these two instruments. Findings will be disseminated to the district school board and local schools.

Description

The SSAT–I is an untimed test containing a section indirectly measuring writing ability through multiple-choice items addressing specific skills such as punctuation, capitalization, spelling, and grammar. The Production Writing Assessment is a direct measurement of writing performance requiring students to write an essay on a given topic within one class period.

For this report, two procedures were used in the Production Writing Assessment. One involved holistic scoring, the other analytical scoring. Holistic scores for the 3386 papers in the sample provided an overview of students' general writing ability. The randomly selected essays were independently rated by professional readers on a scale of 1 (low) to 4 (high). Final scores were the sum of the two scores. Second, randomly selected subsets of each holistic score group in the sample were independently and analytically scored to obtain skill-specific information regarding students' ability to apply the SSAT–I skills.

To estimate the reliability/percent agreement of the two analytical raters, the analytical scores for two skills from papers of the lowest holistic score group (2-3) were compared. For the two skills,the percents of agreement between the two raters' scores were 99% and 97%. Thus it is apparent that any difference between the SSAT–I and the analytically scored results is not likely due to a lack of consistency/reliability in the analytical scoring procedures. Results of the SSAT–I and the analytical scoring were sorted by the four holistic writing score groups (2-3, 4, 5-6, and 7-8) and then compared to determine if the two measurements of the same skills would yield similar results.

Findings

Overall, students perform less well on the SSAT–I items measuring particular skills than on these same skills in actual writing measured analytically. This difference is greatest among students with low holistic writing scores (2–3). The SSAT writing results are of limited use in diagnosing, on a skill-by-skill basis, the type and frequency of errors a below-average student will make in actual writing.

More specifically, the findings indicate the following:

(1) Results of the SSAT–I agree reasonably closely with those from analytically scored production writing only when the students have fairly high levels of overall writing skill.

(2) For below-average students, scores representing the same skills tend to be lower on the SSAT–I than on the analytically scored papers.

(3) The SSAT–I results tend to penalize students with lower levels of overall writing skill. Moreover, the SSAT–I multiple-choice format may even lead to a misdiagnosis, particularly in subject-verb agreement and irregular verb usage, of these students' major problems.

(4) Such diagnostic inaccuracies and/or insufficiencies and the deficiency label placed on schools when SSAT–I scores in a skill area fall below eighty percent passing may lead to incorrectly focused instruction. The possible result can be to harm rather than help the student, and to misinform the instructor and/or the curriculum dsigner. Quite simply, these results appear (to us) to call into question the usefulness of the diagnostic information arising from SSAT–I writing results, at least at grade 10.

(5) Florida law requires that schools with SSAT–I results below a certain minimum be labeled "deficient." Using the writing portion of the SSAT–I as the criteria for labeling schools deficient is at best debatable because the SSAT appears to be a questionable indicator of the actual writing ability of low-level students. Further, because of the nature of writing, exclusive use of the multiple-choice format writing portion does not appear to be an adequate assessment for the measurement of writing skills.

Conclusions

From a somewhat speculative point of view, it is thought that the higher analytical scores arise, at least in part, from students' knowing they do not know, and therefore avoiding words or phrases of which they are uncertain, and substituting structures less likely to be wrong.

In any case, assessing writing skills indirectly through objective measures yields one set of results while direct assessment by means of production writing may yield another. That each has its advantages and disadvantages suggests that the exclusive use of one or the other may lead to misinformation regarding the level of students' writing skills.

A COMPARISON OF 1987 RESULTS OF SSAT-I WRITING AND PRODUCTION WRITING ASSESSMENT

PURPOSE OF STUDY

The purpose of this study is to determine similarities and/or differences between results of the March 1987 SSAT-I Writing and the February 1987 Production Writing Assessment, both administered to Dade County tenth graders. This study is an attempt to ascertain whether student performance on an objective assessment of writing skills matches that on a direct assessment of student production writing.

DESCRIPTION OF ASSESSMENTS UNDER COMPARISON

The State Student Assessment Test Part I is an untimed instrument that contains a section indirectly measuring writing ability through multiple-choice items addressing ten specific writing skills. Among the skills included are punctuation, capitalization, spelling, and grammar. (Refer to Appendix A for a complete listing of the SSAT-I writing skills addressed on the March 1987 test.) Scoring is on a pass-fail basis determined by the criteria of 80% correct (four of five items) for each of nine skills and 70% correct (seven of ten) for the tenth (spelling). Results of the test are used for the initial assessment of students' basic skills (to diagnose whether remediation is needed) and for the determination of schools' deficient/nondeficient status.

The Production Writing Assessment was administered in an attempt to gain an impression of the impact on students of the state-mandated Writing Enhancement Program. It is a direct measurement of writing performance requiring students to write one essay as a response to a given topic (prompt) within one class period. Test directions suggested activities for prewriting and revising/proofreading, and offered advice for budgeting time.

For purposes of the current report, two procedures were used in the Production Writing Assessment. One involved holistic scoring; the other, analytical scoring. First, the holistic scoring method provided an overview of students' general writing ability. In this procedure, professional readers holistically judged each essay's overall effectiveness regarding focus, organization, sentence structure, vocabulary, and mechanics. Randomly selected essays were independently rated by two readers on a scale of 1 (low) to 4 (high). Final scores were the sum of the two readers' scores, from 2 to 8. (Appendix B contains the holistic scoring guide developed by Dade County for use by the scoring company, and examples of students' essays.)

Second, the analytical scoring method was employed to obtain skill-specific information regarding students' ability to apply the nine recently tested SSAT-I writing skills. In this procedure, two local readers independently scored randomly sampled subsets of essays from each of four holistic score groups. Specific skills were isolated and scored individually. In this case, readers scored students' use of SSAT-I skills addressing punctuation, capitalization, spelling, and grammar.

Results of the SSAT-I and the analytical scoring were sorted by four holistic writing score groups and then compared to determine if the two measurements of the same skills would yield similar results.

OUTCOMES EXPECTED OF THE STUDY

Student performance on the objective test (SSAT) was expected to be better than, or at least equivalent to, performance on the analytically scored Production Writing Assessment (PWA) essays for the following reasons:

1. Multiple-choice recognition items are generally thought to be easier than production writing items.

2. Each section of an objective test focuses on a single specific skill, unlike production writing which entails thought development as well as multiple-skill application.

3. The SSAT is untimed, freeing students from the pressure of time constraints. The PWA was administered in one class period, requiring students to read an "unfamiliar" prompt, to plan and write a response to it, and to revise the essay that resulted.

4. According to research done by Hembree and reported in the Journal for Research in Mathematics Education (1987, Vol. 18, No. 3, 197-214), comparisons of multiple-choice and open-response test formats show different results. When answers for both formats were scored right or wrong without partial credit, the multiple-choice format generally yielded higher percentage correct results.

PROCEDURES OF THE STUDY

The classic study of the measurement of writing ability was conducted by Fred I. Godshalk and his coworkers*, the focus of which was to investigate the relative validity of different approaches to the measurement of English composition skills. Its findings indicate that multiple-choice items tend to be somewhat predictive of students' writing skills, but that the best measure of such skills is a combination of multiple-choice items and production writing.

---

*According to Godshalk in The Measurement of Writing Ability (College Entrance Examination Board, 1966), while objective measures of writing skills may be somewhat predictive of levels of students' actual writing performance, the most efficient predictor of a reliable direct measure of writing ability is one which includes essay questions or interlinear exercises in combination with objective questions. In an interlinear exercise, a student is required to find errors in sentence structure and/or grammar within poorly written passages, and then to supply more appropriate forms to correct those deficiencies. Students write their corrections between the lines of the passage, hence the term "interlinear". This is not to be confused with the selection/multiple-choice format found in, for example, the Stanford Achievement Test.

In the current study, the holistic scoring method was employed to gain an overall impression of students' writing ability. However, in order not to overlook specific characteristics of good writing, an analytical scoring guide or rubric was developed by local language arts teachers and other professionals. It is a five-point set of standards addressing focus, organization, sentence structure, vocabulary, and mechanics. This rubric was used to guide the identification of essays as "anchor" papers, those essays which, by consensus, typify each score. This procedure is contrary to traditional holistic scoring. That is, in a true holistic procedure, raters have no predetermined set of standards guiding their score assignments. Instead, their rating is based only on an overall impression of the paper from excellent to poor, relative to all other papers in the study.

However, using the analytical guide to provide a score that reflects only an overall impression results in a combination of the analytical and holistic methods. That is, it requires readers to judge an essay's merits based on pre-established analytical criteria, while also asking them to provide a score for the essay based on its holistic impression only. The committee used the analytical-based rubric to identify papers that typified the characteristics of the score groups they represented. The rating assigned to each anchor paper was the consensus of at least three committee members. Two anchor papers for each score were sent to the North Carolina scoring facility, along with the scoring rubric, to illustrate the standards that the raters were to maintain in making their judgements. Raters' judgements were then subject to standards set by language arts professionals, and were not based on a paper's relative merits. The result was that a modified-holistic scoring method was employed. This procedure, combining the elements of the analytical and holistic scoring methods, provides results that are less subject to the "relativistic" criticism that pure holistic scoring might elicit.

The holistic scores used herein provided a picture of students' overall writing ability and divided their essays into seven score groups from low (2) to high (8). Then, to simplify data analysis, these seven sets of scores were reduced to four: those with holistic scores of 2-3 (107 students), 4 (1739 students), 5-6 (1218 students), and 7-8 (322 students). Then, because holistic scores do not provide skill-specific information, random subsets of papers from the four score groups were analytically scored with regard to the writing skills assessed on the 1987 SSAT-I.

SSAT-I writing data were also sorted by the same low-to-high holistic writing score groups of 2-3, 4, 5-6, and 7-8. Passing/failing percentages were reported by holistic score group for each of the ten writing skills assessed in the SSAT-I. Because one of the skills (B 18, Generate headings for groups of words or phrases) did not apply to production writing, only the remaining nine skills were analyzed in students' essays.

Analytically scoring the essays involved isolating the nine writing skills addressed by the SSAT-I and scoring them individually. According to Spandel and Stiggins in Direct Measures of Writing Skills: Issues and Applications, 1981, analytical scoring is appropriate if the intended outcome is to measure and report students' ability to deal with specific conventions of writing, such as punctuation, capitalization, spelling, and grammar.

Initially, fifteen essays from each writing score group were randomly drawn for analysis, but to corroborate the initial findings in the lowest score group (i.e., holistic scores of 2-3), ten additional papers were randomly drawn for that group only. Using the analytical method of scoring essays explained earlier, two language arts specialists independently rated each essay on nine separate skills.

In order to estimate the reliability/percent agreement of the two analytical scorers, the twenty-five papers from the lowest holistic scoring group were used. This group represents a worst case analysis because it is the group for which there is greatest disagreement between the SSAT and the Production Writing results (see "Results of the Study" later presented), and it is the group for which the students' production writing error rates were the highest. For this group, two of the nine assessed skills were randomly selected to make the comparisons between the two independent scorers' results. The two skills selected were spelling and subject/verb agreement. Because of the time and effort involved in calculating scorers' agreement, thirteen of the twenty-five papers were randomly selected to assess agreement in the skill of spelling, and the remaining twelve were used to assess scorers' agreement in the skill of subject/verb agreement.

Exclusive of the words "a", "an", and "the", the number of words identified as misspelled was calculated for each of the two scorers, and the error rates were subsequently converted to percentages indicating the number of misspelled words per hundred. On average, the difference between the error rates for the two scorers was slightly less than one percent; that is, for 99 out of 100 words identified as being correctly/incorrectly spelled, they agreed on slightly more than 99.

Regarding subject/verb agreement, the second randomly selected skill, the percentages of agreement were slightly lower, but still very high. In this particular case, the average percent of agreement between the two scorers' results, across the twelve papers, was 97%.

Based on these two findings, it is apparent that the scoring done by the two independent analytical readers was quite reliable, and that the difference between the SSAT and the analytically scored results cannot be due to a lack of consistency/reliability in the analytical scoring procedures.

In order to develop a basis for calculating a common type of error rate for the SSAT and the analytical scores of the production writing, the SSAT pass/fail criteria were applied to the analytical scores. That is, eighty percent accuracy was required to pass each skill except spelling (G 58) for which seventy percent accuracy was required on the SSAT. However, in the analytic scoring of the essays, the criteria for spelling mastery was stricter than for the SSAT. There the error rate was based on (1) the holistic writing score itself, because as those scores increased, so did expectations of student achievement; and (2) the consideration that, in production writing, students are more likely to use words they know how to spell. As a result of these considerations, incorrectly spelled words could not exceed seven, five, four, and three percent for writing score groups 2-3, 4, 5-6, and 7-8 respectively.

RESULTS OF THE STUDY

Overall, as the chart on page 6 illustrates, students perform less well on the SSAT items measuring particular skills than on these same skills in actual writing. Moreover, this difference is greatest when students have low holistic writing scores (2-3). These students made ten percent or higher analytical scores on five of the nine skills (i.e., those measured on the multiple-choice SSAT and scored analytically on the written passages), and they made significantly lower scores on none of the nine skills. Only for the group earning the highest holistic scores (7-8) did the results of the two procedures yield essentially equivalent results. Even here, however, there was some disagreement on two of the nine skills; in one case, SSAT scores were higher and in the other, the analytical results were higher.

Somewhat more specifically, in the three skills focusing on subject-verb agreement, use of irregular verbs, and spelling words with suffixes, the SSAT results tend to differ substantially for the low holistic score group. SSAT results here were twenty or more percentage points lower than the results from the analytical scoring. A more detailed discussion for the low holistic score group follows.

SSAT-I, Skill A 10 - MAKE SUBJECTS AND VERBS AGREE
Only 38% of the students in the lowest holistic score group passed this SSAT skill by identifying the correct form of subject or verb in at least four of five given items, but 56% of the essays scored analytically contained no errors at all in this skill. Another 36% of the essays had no more than twenty percent errors; therefore, 92% of the 2-3 students demonstrated "mastery" of this skill in actual writing practice, a difference of fifty-four percentage points over the SSAT results.

SSAT-I, Skill A 11 - USE APPROPRIATE FORMS OF COMMON IRREGULAR VERBS
On the SSAT, this skill was passed by 43% of the students in the lowest holistic score group. Of the twenty-five essays studied from this group, 52% were completely free of errors in the use of irregular verbs, and another 24% had no more than twenty percent errors. Thus, 76% of these students demonstrated this skill successfully in actual writing, 33% more than on the SSAT.

SSAT-I, Skill G 61 - APPLY GENERALIZATIONS FOR ADDING COMMON SUFFIXES
On the SSAT, 66% of the students in the lowest holistic score group passed this skill, but, by applying the SSAT error rate to the analytical scores, all in this group "passed" in actual writing.

(For a more thorough presentation of the skill-by-skill comparison for the four groups, see Appendix C.)

Considering the overall results for the study, there was only low-to-moderate agreement between the SSAT-I writing scores and the PWA holistic scores. The (Pearson) correlation was .42 across the nearly 3200 students. On average, students with low holistic scores tended also to have lower SSAT multiple-choice scores, but the skill-by-skill comparisons provided herein indicate quite strongly that skill-specific SSAT writing results are of limited use in

PERCENT PASSING WRITING SKILLS IN SSAT-I AND PRODUCTION WRITING ASSESSMENT

HOLISTIC WRITING SCORE GROUPS

SKILLS	Low 2-3		Average 4		High Average 5-6		High 7-8	
	SSAT	PWA	SSAT	PWA	SSAT	PWA	SSAT	PWA
A 10 Subj-verb Agreement	38 **	92	60 **	100	81 **	100	93	100
A 11 Irregular Verbs	43 **	76	75 **	87	89	93	97	100
B 20 Organize Topic	65	64	81 **	100	93	93	97	100
G 58 Spelling	68 **	80	85	87	95	100	99	100
G 61 Spelling Suffixes	66 **	100	81 **	100	90 **	100	98	100
H 68 Comma in Dir. Add.	56	(not used)	74	(not used)	91	(not used)	95	(not used)
H 69 Apostrophe Poss. Nouns	32	29 (28% used)	47 *	29 (47% used)	61	63 (53% used)	73 *	33 (60% used)
I 74 Caps in Titles	16	(not used)	37	(not used)	53 **	80 (33% used)	73 **	100 (27% used)
I 75 Caps on N/Adj	44 **	59	73 *	50	87 *	67	92	85 (87% used)
# of randomly drawn papers	25		15		15		15	

	2-3	4	5-6	7-8	Total of 9 skills and levels
**SSAT 10%+ < Production Writing	5	4	3	1	13
*SSAT 10%+ > Production Writing	1	2	1	1	4

Usage rates in PWA below 25% are shown as "not used".

indicating the type and frequency of errors a student will make in actual writing. That is, the SSAT writing results can hardly be considered diagnostic, on a skill-by-skill basis, for below-average students.

From a measurement point of view, the weakness in SSAT "diagnoses" is quite understandable, even predictable, and this for at least two reasons. First, there are simply not enough questions on any skill for the results, i.e., the diagnosis, to be reliable. Secondly, the questions used to "measure" a particular skill more than occasionally appear to have questionable content validity. At times this is due to the nature of the skill when placed into a multiple-choice assessment format, e.g., in Skills G 58 and G 61, where students' spelling ability is assessed by identifying the correctly spelled forms from a controlled vocabulary list, rather than by correctly spelling words used in composition. At other times, it appears that one or more specific questions are at best rather obtuse measures of the skill, particularly at a minimum-performance level. To illustrate, in Skill H 69, students are to identify a correct possessive form using an apostrophe and "s" for nouns not ending in "s", as in, for example, "wharf's".

For still another reason, the two sets of scores (one derived from analysis of written passages and the other from skill-specific multiple-choice items) may well differ. In actual writing, the student may know he does not know, and therefore avoid using the word or phrase in question, substituting in its place another that he does know. This knowledge/skill is surely important in writing, although it can never be measured in a multiple-choice format - a problem somewhat analogous to (but more critical than) measuring the skill of "estimation" in mathematics.

CONCLUSIONS OF THE STUDY

Effort must be made to consider what are regarded as adequate levels of writing ability. For example, what is the effect of correct subject-verb agreement on a student's overall writing ability? As has been mentioned, 92% of the students in the lowest holistic score group have demonstrated that they can write sentences in which subjects and verbs agree, and this with few errors. However, does this mean that they are good or even moderately good writers? What must be considered is that students of low ability might know that they do not know how to deal with the more difficult or complex structures. Therefore, by avoiding, for example, the use of prepositional phrases between subject and verb, students can then write with minimal errors in this skill. Yet the same group of students can and do respond with differing levels of skill on multiple-choice items requiring an identification of the correct structure, perhaps from a list of controlled difficulty. This difference calls into question what one means by <u>adequate</u> levels of writing ability.

The types of writing range from that generated by students who use simple structures incorrectly or correctly to that of students who use complex structures incorrectly or correctly. Where "adequate writing" falls along the continuum between these extremes is not a question that can be easily resolved, but it is one which must be recognized. Assessing writing skills by means of objective measures yields one set of results while direct assessment through production writing may yield another. That each has its advantages and disadvantages suggests that the exclusive use of one or the other may lead to misinformation regarding the level of students' writing skills.

Overall, then, the data provided herein indicate the following:

(1) Results of the SSAT-I agree reasonably closely with those from analytically scored production writing only when the students have fairly high levels of overall writing skill.

(2) For below-average students, scores representing the same skills tend to be lower on the SSAT-I than on the analytically scored papers.

(3) The SSAT-I results tend to penalize students with lower levels of overall writing skill. Moreover, the SSAT multiple-choice format may even lead to a misdiagnosis, particularly in verb usage, of these students' major problems.

(4) Such diagnostic inaccuracies and/or insufficiencies, and the deficiency label placed on schools when SSAT-I scores in a skill area fall below eighty percent passing may lead to incorrectly focused instruction. The possible result can be to harm rather than help the student, and to misinform the instructor and/or the curriculum designer. Quite simply, these results appear (to us) to call into question the usefulness of the diagnostic information arising from SSAT-I writing results, at least at grade 10.

(5) Using the writing portion of the SSAT-I as part of the criteria for labeling schools deficient is at best debatable because the SSAT appears to be a questionable indicator of the actual writing ability of low-level students. Further, because of the nature of writing, exclusive use of the multiple-choice format writing portion does not appear to be an adequate assessment for the measurement of writing skills.

APPENDIX A

SSAT-I WRITING SKILLS ASSESSED IN 1987

SSAT-I WRITING SKILLS ASSESSED IN 1987

Standard/Skill	Criterion	Skill Statement
		Sentence Composition
A 10	4/5	Make subjects and verbs agree.
A 11	4/5	Use appropriate forms of common irregular verbs.
		Organization
B 18	4/5	Generate headings for groups of words or phrases.
B 20	4/5	Organize information related to a single topic.
		Spelling, Punctuation, Capitalization
G 58	7/10	Spell words needed in writing through grade ten.
G 61	4/5	Apply generalizations for adding common suffixes.
H 68	4/5	Use the comma to set off proper names in direct address.
H 69	4/5	Use apostrophe and "s" for possessive nouns not ending in "s."
I 74	4/5	Capitalize appropriate words in titles.
I 75	4/5	Capitalize proper nouns and proper adjectives.

APPENDIX B

PRODUCTION WRITING ASSESSMENT

PART I: HOLISTIC SCORING GUIDE

HOLISTIC SCORING GUIDE

4 POINT

An excellent paper has all or most of the following:
- a clear focus which is appropriate to the prompt
- good organization including an introduction, details, and a conclusion
- effective sentence structure, appropriate to the topic and varied in pattern
- specific and vivid vocabulary that shows insight into the topic
- few or no errors in conventions of standard American English

3 POINT

A good paper has all or most of the following:
- a focus which is generally clear and appropriate to the prompt
- good organization which may, however, lack a clear introduction, sufficient details, or a conclusion
- effective and appropriate sentence structure in most sentences
- appropriate vocabulary that shows some insight into the topic
- some errors in the conventions of standard American English

2 POINT

An adequate paper has all or most of the following:
- a focus which is appropriate but in many parts unclear
- unclear organization in much of the composition
- generally adequate sentence structure which may contain awkward constructions
- many errors in the conventions of standard American English

1 POINT

An inadequate paper has all or most of the following:
- a focus which is inappropriate and generally unclear
- no recognizable organization
- inadequate sentence structure that detracts from the meaning
- vague or dull vocabulary
- many serious errors in the use of the conventions of standard American English

0 POINT

This paper is blank, illegible, written in a foreign language, has insufficient amount of writing to evaluate, or is totally inappropriate to the task.

APPENDIX B

PRODUCTION WRITING ASSESSMENT

PART II: EXAMPLES OF HOLISTICALLY SCORED STUDENTS' ESSAYS FROM THE 1986 PRODUCTION WRITING TEST

1986 PRODUCTION WRITING TEST

PROMPT: Consider all of the teachers you have had in school. Identify one that you consider unforgettable. Describe that person so the reader will know why he or she is so memorable to you. Include incidents or memories to show why that teacher is unforgettable.

HOLISTIC SCORE: 4

PROMPT NUMBER: 13

~~[illegible]~~

Five years ago, in my fifth grade year, I was taught by a teacher that I consider unforgettable. Her name was Mrs. ———. She had brownish-grey hair and she was very plump. She was always smiling and her eyes were always bright. To me she looked like the ideal Mrs. Santa Claus.

One reason I will never forget her is that she helped me a great deal in learning English. She stayed after school one day for one hour and a half helping me understand ~~[illegible]~~ the very complicated structures of a sentence; which afterwards seems very simple ~~[illegible]~~.

Mrs. ——— was from Belgium and she called me by saying "Lee." I asked her why once and she

~~She~~ replied, "In Belgium everybody is 'love,' so I say it automatically. I liked it ~~when~~. She called me "love." It made me feel like I was her close friend, and not just her student.

I still can remember having long conversations with her after school. She would be grading papers and I would clean her chalkboard, or something of that nature, and we would talk about everything. She used to tell me about her son who was trying to become a singer; ~~I~~ she even gave me one of his records. I also remember how I talked to her about "growing up" problems. I think she knew something about everything.

Mrs. ~~[illegible]~~ was, by far, my favorite teacher. ~~I~~ I'll had many wonderful days in her classroom. I will never forget her. (I wonder, though, if she remembers me.)

Prompt Number: 13

HOLISTIC SCORE: 4

PROMPT NUMBER: 13

An Unforgettable Teacher

My most memorable teacher is my fourth grade teacher, Mrs. ——. She is very special to me, because she helped me to get through one of the worst periods in my life, which was my parents divorce.

I was just a seven year old child, who was ignorant to the realities of life. When I was sent to school I didn't go, I found myself a hiding spot, and waited until my mother had left for work, then I would emerge. I really hated my mother, because she was the one who had left my father, but I didn't know why. He was such a wonderful man. I even ran away from home once, hoping that they would see what it was doing to me, and re-unite. But as much as I tried

it wouldn't happen.

One day, my mother was called to my school, for a parent-teachers conference, where she found out that I had been cutting classes. My mother explained to my teacher that I was going through some difficulties at the time. Mrs ———, my teacher was very sympathic with me from then on. She became my confidante, we talked during lunch break, and she always knew what to say to make my world a little brighter. I even started to think of her as a mother, since I resented my mother at that time.

Mrs. ——— never tried to make me hate my mother, she did the opposite. She always kept telling me that my mom must have had a good reason for divorcing my father. She often told me that life was very complicated and one day I would understand. Mrs. ——— and I grew very close because of the situation I was in, atleast someth-

Prompt Number: 13

good came out of it.

Three years past by, and I had to leave elementary school, which meant that I couldn't see Mrs. ——— as much as I wanted to. I was very melancholy at the prospect of leaving her, but she said I could come and see her whenever my heart desired.

I had grown older and was beginning to understand a little more about life, but although I found out why my parents divorced, It didn't stop me from hoping, and praying that someday they would remarry and we would be a happy family again. I still am praying today. In conclusion, I would like to say that Mrs ——— was, is a very important and unforgettable person in my life.

Prompt Number: 13

1986 PRODUCTION WRITING TEST

HOLISTIC SCORE: 3

PROMPT NUMBER: 13

Out of all the teachers I have had while attending schools, the most memorable was Mr. ——— Mr. ——— was a seventh grade mathematics teacher. I enjoyed having him as a teacher because he really made learning fun. On Mondays he would allow us to finished unfinished homework, then, he would assign us our classwork. Usually if no one understood it he would help us by making a game out of a very hard Problem. Tuesdays through Thursdays were good days also, He allowed us to sit near a friend to help one another and if we were good we could play math games. Fridays were the best days we ever had. We would first take our test (which we usually got A's and B's on) then we'd play a math game by ourself until everyone was done, then, we would

take turns asking mystery questions until some one guessed them or until we would get tired. ~~Then~~ Later we would tell jokes. When we had parties mister "O" would always furnish the food and we would furnish the music. As the year finally went by and everyone went to eighth grade we always would drop by every now and then to see our favorite teacher. Although mister "O" played lots of games, they were educational (math) games he was the only teacher I remember who ever made learning fun, he taught us alot about life, fun and school. If I had a chance to go back to the Philippines my first stop would be to stop in room 108 to see mister ——— and thank him for the memories and the good times we had in room 108, but I know I'll never see him again so I thank him by doing my best in school, the very best that I can possibly ever do. I am sure that most

Prompt Number: 13

teachers are capable of making learning fun and they usually try. But the never suced because they don't know that the key is to be young at heart and be willing to have a little fun. If most teachers were like this the students would more than likely make good grades.

Prompt Number: 13

1986 PRODUCTION WRITING TEST

HOLISTIC SCORE: 3

PROMPT NUMBER: 73

A teacher most memorable to me would be Mrs. ——— my old English teacher from my old school of ——— Jr. High. She was a woman of the up-most characteristics. It was her way of teaching that brought new patterns of thought through my mind and all of the rest of the students. She would teach us with the greatest patience a person could ever have. When she saw that I or any one else did not understand what she was teaching about she would teach it again and again until it was engraved within our heads. I will never forget her and all the memories and incidents I experienced from that time.

The best memories from her class was all the wise-cracks she would give us when didn't do our homework or write required assignments. Sometimes she would embarrass us by putting someone in the corner when he or she talked to much which made us

ful like little children. Thus, her ways were good and many of the memories I have will never fade away.

Prompt Number: 13

1986 PRODUCTION WRITING TEST

A teacher most memorable to me would be Mrs. —— from my old school of —— Jr. High. She was a woman of the upmost characteristics. She was bright, intelligent and very understanding. Her way of teaching brought new patterns of thought through my mind and the rest of the students. She would teach us with the greatest patience a person could ever have. When she saw that we did not understand she would teach it again and again until it was ingrained within our heads. I will never forget her and all the memories and incidents from that time.

Prompt Number: 13

1986 PRODUCTION WRITING TEST

HOLISTIC SCORE: 2

PROMPT NUMBER: 13

The most unforgettable teacher that I have had is Mr. ——— my 7th, 8th and 9th grade band teacher. My memories of him are that he would treat people different than other teachers. He wasn't mean he just would have the wrong idea he wouldn't see things from a kids points of view.

If something occured he would blame everything on the whole class. Suppose one person did something wrong and he got sent down to the office Mr. ——— would compare us to him. He would say that it's all these kids you're all degenerates etc this generation. He said also that it was better when he was a kid back in Italy.

I never paid too much attention to him I just did things I had to do. When I was in 7th grade he was nice then in 8th he was awful. Then in the biginning of 9th he was even worse, but in the second qtrm of 9th grade a metamorphises occured he became a real nice hearted person.

33

He had really changed. he would treat everybody nice and kind. he would give us days off from playing music. I wouldn't call it music because the band didn't know how to play. Mr —— tried his best, that's why I think of him as an unforgettable teacher.

Prompt Number: 13

1986 PRODUCTION WRITING TEST

HOLISTIC SCORE: 2

PROMPT NUMBER: 13

Mrs ——— is memorable to me in many different ways. I had her in 4th grade for math and I learned alot in her class. And then when math was over we played the flutes. Our class learned many different song. (It was alot of fun.)

Near the end of the year we had our class party and boy was that some fun. We made Jewelry Boxes for our parents. After we got done doing that we ate cookies, cake, and milk. Alot of mothers were there to serve us. Because

we would make a mess

At the end of the year, we had Bus talent Show and Mrs. —— had her 4th grade class play the flutes and She played along with us. When we were done the principle got up on the stage and he said that we were real good. So that made Mrs. —— feel good and us too. The principle also asked us to play this song called "Lokey evening" (which was Mrs. Sievers favorite song) So we played it. And then we got off of the stage.

Mrs. —— is one of my most memorable teacher. I still go and see her and She

Prompt Number: 13

always say ——— you
Still play your flute?

Prompt Number: 13

1986 PRODUCTION WRITING TEST

HOLISTIC SCORE: 1

PROMPT NUMBER: 13

My Band teacher who I feel unfortunate. On this grading period he gave me a C instead of an A. just because I'm not going to state contest and I don't think that's fair because we all had to try out for the contest first and I've happen not to make it.

1986 PRODUCTION WRITING TEST

HOLISTIC SCORE: 1

PROMPT NUMBER: 13

Considering all of the teachers I have had in school, I consider that one is unforgettable and she a Science teacher, she is not unforgettable only by me, by everybody. Everybody loves her, because she is a sweet, intelligent, person she is a very good teacher. I think she is the best teacher in the world.

This teacher helps everybody in every problem the person have. Sometimes I have a problem, and I tell her and she is always telling me what to do in that moment. She is a good teacher she explains good what she is teaching.

I think she is a wonderful person, and also I think that she is very happy of being a teacher because she knows that every student that she have loves her. I'm glad I have her for one class, and I only wish that every teacher could be like her. Everybody knows that she is the best teacher, and every students loves her for everything she does for us. She teaches is ____ High, her name is Mrs ____, and I'm proud of having her as a class, because she's the best teacher in the world.

Prompt Number: 13

APPENDIX C

RESULTS OF ANALYTICAL SCORING

COMPARISONS OF RESULTS FOR SKILLS WHERE GREATEST DATA DIFFERENCES OCCUR

COMPARISONS OF RESULTS FOR SKILLS WHERE GREATEST DATA DIFFERENCES OCCUR

A 10 - MAKE SUBJECTS AND VERBS AGREE

This skill was passed on the SSAT by 38% of the students in holistic writing score group 2-3, 60% of score group 4, 81% of group 5-6, and 93% of group 7-8. But in actual writing performance, students in all groups "mastered" this skill much more frequently than one might expect from the SSAT scores.

The greatest difference between analytical results and the SSAT scores occurs among students with holistic scores of 2-3. Only 38% of these students passed this SSAT skill by identifying the correct form of subject or verb in at least four of five items, but in actual writing, 56% made no errors at all in this skill. Another 36% of the papers contained up to twenty percent errors, for a total of 92% of the 2-3 students demonstrating "mastery" of this skill in actual writing practice, a difference of fifty-four percentage points more than the SSAT results.

Of the students whose holistic writing score is 4, 60% passed the subject-verb agreement skill on the SSAT. However, in the fifteen essays studied from this group, 87% had no errors in subject-verb agreement. Moreover, no student had more than ten percent errors. Thus, there is a difference of forty percentage points between the analytical scores and the SSAT results for this skill.

Among the remaining two score groups, 5-6 and 7-8, all students "mastered" this skill in actual writing; whereas, on the SSAT, 81% of the former and 93% of the latter passed this skill.

What explanation can be suggested for these occurrences? They differ from the expected outcome, which says that poor performance on objective test items should reinforce/be reinforced by equally poor performance on production items. Comparing the subject-verb forms contained in the SSAT items with the subject-verb forms supplied by the essay writers might be helpful.* Among the five SSAT items were two sentences having plural subjects needing simple plural verbs. For these, students had to select the correct subject form for one and the correct verb form for the other. A third item was a sentence with a singular subject followed by a prepositional phrase. For this, students were required to select a simple singular verb. In the remaining two items, one subject was compound, the other was plural, preceded by a compound adjective. Students were required to select similar answers, both plural verb phrases containing an auxiliary verb and the past participle form of an irregular verb (even though the skill does not specify students' knowledge of irregular verb forms).

In their essays, students used many simple plural subjects and simple plural verbs similar to the first two aforementioned SSAT items (i.e., plural subjects needing simple plural verbs). However, only a few essays contained the compound subject/plural verb form, and only in the 7-8 group did essays contain a subject and verb interrupted by a prepositional phrase. The most frequently used forms consisted of a singular or plural simple subject with a

---

*To maintain test security, actual SSAT-I test items are not included in this report. Their substance, however, is connoted for the sake of clarity.

singular or plural simple verb. Verb phrases, when employed, usually consisted of was-were or has-have and a present or past participle.

Students tend to select forms with which they are familiar and comfortable. This accounts for the high levels of performance across all holistic score groups in student-generated subject-verb agreement. The frequency of simple subject/simple verb usage suggests that these forms are the most familiar to students, and this is the probable reason they are used more frequently than others. However, items of this type appeared only two or three times (of the five) on the SSAT. Even if students answered these two or three correctly, mere guessing at the remaining items might not have been enough to correctly answer the four or five items required to demonstrate mastery of the skill.

A 11 - USE APPROPRIATE FORMS OF COMMON IRREGULAR VERBS

On the SSAT, students demonstrated this skill by selecting one of four verb forms to complete a given sentence. This SSAT skill was passed by 43%, 75%, 89%, and 97% of holistic writing score groups 2-3, 4, 5-6, and 7-8 respectively. Comparing SSAT results to students' actual writing practices shows that while little difference exists between the two scores among students in the three highest holistic score groups, again a difference is found in the 2-3 group. All papers studied for this report employed some forms of irregular verbs. Of the twenty-five essays from the low holistic score group, 52% were free of this error, and another 24% had less than the maximum twenty percent error rate allowed by the SSAT. Thus, 76% of these students demonstrated this skill successfully in actual writing, one third more than on the SSAT. In the next higher writing score group, 4, 53% of the papers contained no errors, and another third had up to sixteen percent errors, for a total of 87% demonstrating correct use of appropriate irregular verbs and thus "passing" this skill; whereas, 75% passed this portion of the SSAT.

Four of the five irregular verbs featured in the SSAT items were used by students in their essays, but many other irregular verbs also appeared. Most frequently used were "began," "knew," "took," "left," "felt," and "thought," as well as forms of "be," "go," and "do." The essays in the 7-8 group were more likely than the others to contain less common irregular verbs, such as "stung," "held," "wound," "swam," and "loosen."

G 58 - SPELL WORDS NEEDED IN WRITING THROUGH GRADE 10

In this skill also, the 2-3 papers contained fewer overall errors than the SSAT score for the same group would suggest. The SSAT allows a maximum error rate of thirty percent for this skill, but in the analytical scoring, the maximum was seven percent for students in the 2-3 group. This was done because students are likely to use only those words that they think they can spell correctly. The smaller allowance for errors somewhat compensates for the difference between the two kinds of "items," that of identifying ten correctly spelled words taken from a master list (Florida Lists for Assessment of Spelling, or FLAS), and that of supplying correctly spelled words from the student's own "master list." Even with this lower error rate applied to their essays, 80% of the students in the 2-3 group "passed" this skill. However, it must be noted that thirty-nine percent of the 209 words misspelled by these students are FLAS words from grades three, five, eight, and eleven. Only two students had no errors in FLAS words, and only seven students had five or more errors in FLAS words.

G 61 - APPLY GENERALIZATIONS FOR ADDING COMMON SUFFIXES

This skill requires students to identify the correct spelling of a word formed by adding a given suffix to a given word. For example, a typical item might be the following: If the suffix -ly is added to the word happy, which answer is the correct spelling of the new word? The options might include spelling variations such as "happly," "happyly," happilly," as well as "happily."

With that in mind, the essays were studied for correct spellings of words with suffixes. Using the SSAT error rate of twenty percent as the criterion resulted in all students "passing" with 100% accuracy. This included the lowest group, of which only 66% passed the same skill on the SSAT. Among essays in writing score group 4, analytical scores for this skill were nineteen percentage points higher than on the SSAT. Essays in writing score groups 5-6 and 7-8 earned, respectively, ten and two percentage points more than for the SSAT.

Most common in the 25 essays of students in the low holistic score group were the correctly added suffixes indicating plurals (19 essays) and present or past participles (22 and 16 essays respectively). Each form was misspelled three times. The -ly, -er, -y, and -ion suffixes were found next most frequently (in 8, 5, 4, and 3 essays each) with two misspellings occurring only in the -ly form. Several other suffixes appeared without errors in only one or two essays each, such as -est, -ful, -ous, and -en.

I 75 - CAPITALIZE PROPER NOUNS AND PROPER ADJECTIVES

On this skill, the group of students with holistic scores of 2-3 performed better in actual writing than on the SSAT, although the difference between the two measures is smaller here than for any other skill. However, for all other groups of students, analytical scores here were lower than were the SSAT scores.

Recent SSAT-I items used to measure this skill contained context clues which may have led students to select the correct response. In these items, students were to identify that capitalization was needed in the following cases:

1. The name of a holiday, for which the test stem contained the word "holiday";
2. The name of a corporation which the stem implied to be a place of business;
3. A three-word name of a school, which was not likely to be easily confused with the common noun "school" preceded by two common adjectives because the first two words were capped;
4. The name of a mythological figure referenced by the word "goddess" in the stem; and
5. The name of a language, for which the test stem contained the name of another language as well as the word "language."

In actual writing, most students used a variety of proper nouns, but few used proper adjectives. Most proper nouns were names of persons, with names of places occurring next most often. For example, in writing score group 5-6, where analytical scores on this skill were down twenty percentage points from the SSAT results, all papers employed this skill, and names of persons appear twice as frequently as names of places.

BUREAU RESPONSE

TO REPORT*

*The following section is a response initiated and developed by the Bureau of Education program staff and is not an official part of the report itself.

December 16, 1987

RESPONSE TO THE FEBRUARY 1987 TENTH GRADE PRODUCTION WRITING ASSESSMENT REPORT

A review of the February 1987 Tenth Grade Production Writing Assessment has been completed by the Bureau of Education staff. As a result of this review and based upon the information contained in the referenced report, the following recommendation is made:

> Since there are implications for the Department of Education, the study should be transmitted to the Assessment, Testing, and Evaluation Section of the Florida Department of Education for its considersation.

This recommendation has been reviewed by Mrs. Zelda Glazer, Dr. Charles Sherwood, Dr. Gloria McPhee, and Mr. Richard O. White, and has the concurrence of this office.

Frank de Varona

The School Board of Dade County, Florida adheres to a policy of nondiscrimination in educational programs/activities and employment and strives affirmatively to provide equal opportunity for all as required by:

Title VI of the Civil Rights Act of 1964– prohibits discrimination on the basis of race, color, religion, or national origin.

Title VII of the Civil Rights Act of 1964, as amended - prohibits discrimination in employment on the basis of race, color, religion, sex, or national origin.

Title IX of the Education Amendments of 1972 - prohibits discrimination on the basis of sex.

Age Discrimination Act of 1967, as amended - prohibits discrimination on the basis of age between 40 and 70.

Section 504 of the Rehbilitation Act of 1973 - prohibits discrimination against the handicapped.

Florida Educational Equity Act - prohibits discrimination on the basis of race, sex, national origin, marital status or handicap against a student or employee.

Veterans are provided re-employment rights in accordance with P.L. 93–508 (Federal) and Section 295.07, Florida Statutes, which also stipulates categorical preferences for employment.